PROCEEDINGS

Strategic Frameworks for
Human Performance and Human Systems Integration

Spotlight on
General Systems Performance Theory and
Nonlinear Causal Resource Analysis

Workshop
Motivated and Convened by the
Human Systems Integration Task Force
Defense Safety Oversight Council
Major General Thomas Travis, Chair HSI TF

December 13-14, 2010
Human Performance Institute
The University of Texas at Arlington
Arlington, TX

All Sessions:
University of Texas at Arlington
Automation and Robotics Research Institute Auditorium
Riverbend Campus, Fort Worth, TX

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 JAN 2011**	2. REPORT TYPE	3. DATES COVERED
4. TITLE AND SUBTITLE **Strategic Frameworks for Human Performance and Human Systems Integration: Spotlight on General Systems Performance Theory and Nonlinear Causal Resource Analysis**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defense Safety Oversight Council,4000 Defense Pentagon, Suite 1E532,Washington,DC,20301-4000**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
A two-day workshop was held at the University of Texas at Arlington to discuss, analyze, and make recommendations on frameworks for human performance and human systems integration usable by the DoD. Twenty-five participants from eighteen organizations listened to nineteen presentations on human performance frameworks and their problems and applications. The presentations and the discussions they sparked are detailed in these proceedings. The group converged on recommendation of General Systems Performance Theory (GSPT) and its derivative, Nonlinear Causal Resource Analysis (NCRA), as composing a framework with potential to address a host of problems faced by the DoD in making the best use of human performance advances in today?s dynamic environment. Specific recommendations to the Defense Safety Oversight Council are contained in the report.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**32**	

Preface

The Defense Safety Oversight Council chartered the Human Systems Integration (HSI) Task Force to improve safety and reduce mishaps in all services with the application of HSI principles and processes. I have been the chair of the HSI Task Force since its inception in 2008.

In 2009 as part of its findings, the HSI Task Force identified the need for a framework in which to organize human performance. This need is driven by increased knowledge in all aspects of human performance coupled with increased demands on humans as decision makers and as system operators.

Technologic advances increasingly make the human not only the weak link in the chain, but often the single point of failure. Nowhere is this more evident than in the military, where a single individual may command a weapon system with the capability to engage many targets with great precision, while receiving in a matter of minutes amounts of information that would have taken weeks or months to compile a few decades ago. Future visions include increased use of autonomous systems; to the casual observer, this would seem to mean less reliance on humans, but ultimately it means the performance of each single human who programs or directs autonomous systems becomes even more critical.

This document represents the dedicated work of twenty five participants from eighteen organizations over two days, attacking the problem of finding an organizing framework for human performance. Participants from Army, Navy, Air Force, Marine Corp, NASA, and civilian organizations contributed. Each of those participants was chosen for experience and expertise in fields touching human performance and HSI. Their objective was to make specific recommendations for use by the Defense Safety Oversight Council on adoption of a framework for human performance for the Department of Defense.

I am pleased to endorse their recommendations, along with summaries of the presentations and debates that led to these recommendations. This is an important step forward for the safe, effective employment of our service members and a critical step in support of Human Performance and HSI programs in all the services.

THOMAS W. TRAVIS
Major General, USAF, MC, CFS
Deputy Surgeon General

Executive Summary

A two-day workshop was held at the University of Texas at Arlington to discuss, analyze, and make recommendations on frameworks for human performance usable by the DoD. Twenty-five participants from eighteen organizations listened to nineteen presentations on human performance frameworks and their problems and applications. The presentations and the discussions they sparked are detailed in the attached proceedings.

By late morning on the second day, the group had converged on recommending exploration of General Systems Performance Theory (GSPT) and its derivative, Nonlinear Causal Resource Analysis (NCRA), as composing a framework with potential to address a host of problems faced by the DoD in making the best use of human performance advances in today's dynamic environment.

The recommendations are summarized as follows:

- There is strong consensus that GSPT and NCRA show great promise as a very capable, evidence-based framework with multiple military applications.

- Correct application to support decision makers has the potential to: improve personnel selection; reduce attrition; optimize training; inform research in science and technology; improve and validate human centered standards for acquisition requirements, development, and engineering [Human Systems Integration (HSI) standards]; improve performance of individuals, teams, and systems; and thereby reduce costs and mishaps.

- Recommendations (further detailed in the conclusions section):

    1. Demonstrate the concept by analyzing existing datasets that will rapidly result in actionable information

    2. Fund NCRA analysis of performance data in selected new studies

    3. Fund maturation of NCRA software

    4. Broaden the understanding of GSPT and NCRA in the safety and human performance communities

    5. Follow up with future task force and human performance community meetings to assess and guide progress

Background

The Human Performance Working Group (HPWG) of the Human Systems Integration (HSI) Task Force (TF) in 2010 identified the lack of a common conceptual framework for characterizing and understanding human performance and human systems integration problems across the services and across different problem domains (e.g., medical vs. non-medical, cognitive vs physical, etc). The HPWG subsequently identified General Systems Performance Theory (GSPT) and its derivative, Nonlinear Causal Resource Analysis (NCRA), as a novel and potentially significant framework for addressing a host of highly relevant application problems within the DoD context. This workshop emerged as the first step to: 1) increase awareness of conceptual framework issues, 2) stimulate critical debate regarding GSPT, NCRA, and related concepts, 3) promote focused discussion throughout relevant DoD communities, and 4) initiate a process to achieve consensus on the characteristics of a framework for HP and HSI.

Format

Participants were presented with information on frameworks and the benefits they can have in support of the various roles that human performance and human systems integration play in DoD. The organizing committee chose GSPT, and its derivative, NCRA, as the reference framework against which any others would be compared. Presentations alternated with facilitated discussion of conceptual frameworks utilized in addressing respective human performance issues and the value and limitations of such frameworks. The agenda placed emphasis on allowing time for detailed discussions.

Proceedings

This is a summary of presentations and discussions. Discussions, including the moderated discussion sessions, are non-attributional, except that when a question and answer format is indicated, the answers, indicated by the notation "A," were given by the relevant speaker. Discussion often occurred during a presentation. This is indicated by the use of "Comment" or "Q" to indicate a different speaker without a separate "discussion" heading. When a paragraph in such a sequence is unattributed, the main speaker has resumed the presentation.

Participants

Name		Branch	Unit
Armstrong, Jonathan		Civ	University of Texas at Arlington, Human Performance Institute
Belland, Kris	CAPT	Navy	Force Surgeon, Naval Air Forces
Brown, Lex	Col	USAF	Director, Human Performance Directorate, 711th Human Performance Wing
Brown, Randy		USAF	Chief Engineer, 711th Human Performance Wing
Chandler, Steven	COL	Army	Chief, Human Dimension Task Force, Requirements Integration Directorate
Cohn, Joseph	LCDR	Navy	Military Deputy, Human and Bioengineered Systems Division, Office of Naval Research
Ensign, Wayne	Dr	Navy	Research Physiologist, Naval Space and Warfare Systems Command
Fiedler, Edna	Dr	Civ	Aerospace Psychology Consultants
Fitts, David		NASA	Chief, Habitability and Human Factors Branch, Johnson Space Center
Francis, Michael	CDR	Navy	Naval Safety Center
Knight, Derek	Col	USAF	Incoming Director, Human Performance Directorate, 711th Human Performance Wing
Kondraske, George	Dr	Civ	University of Texas at Arlington, Human Performance Institute
Kramer, F. Matthew	Dr	Army	Natick Soldier Research, Development & Engineering Center
Lowe, Don		USAF	Security Forces Center, Lackland AFB
Lowry, Cheryl	Lt Col	USAF	Fight Medicine Standards, AF/SG3P
Mann, Benjamin	Dr	Civ	Ayasdi, Inc
Martindale, Valerie	Lt Col	USAF	Plans and Programs Division, 711th Human Performance Wing
Owen, John		Navy	Naval Air Systems Command, Human Systems Integration
Padgett, William	CDR	USMC	Preventive Medicine, HQ USMC
Savage-Knepshield, Pamela	**Dr**	Army	Human Factors Integration Division Chief, Army Research Lab
Sipes, Walter	Dr	NASA	Johnson Space Center
Slyter, Thomas	Col	USAF	Human Factors Division Chief, USAF Safety Center
Statkus, Michael		Army	Natick Soldier Research, Development & Engineering Center
Stewart, Malcolm	Dr	Civ	Director, Presbyterian Hospital of Dallas, Human Performance Lab
Ward, Robert	Dr	Civ	Sport Science Network

13 December 2010

Major General Thomas Travis, USAF, Chairman of the HSI Task Force (TF) under charter to the Defense Safety Oversight Council, made opening remarks by video teleconference from Washington, DC. MGen Travis welcomed participants. He expressed regrets that the Coast Guard representative was unable to attend.

He introduced the basis and expectations for the workshop. Briefly:

1. Advances in knowledge in human performance have been dramatic. At the same time, technologic advances increasingly make the human not only the weak link in the chain, but often the single point of failure.
2. How do we enable the human to take advantage of these advances? How do we select, train, and equip the human to accomplish the mission? How do we use wisely the ability to enhance human performance to result in safer, more effective operations overall?
3. The first task is to understand the problem. What do we mean by a framework? What attributes does a framework need to have? What groups of people will make use of it?
4. George Kondraske's General Systems Performance Theory will serve as the reference framework for the workshop.
5. Others are invited to present information about other frameworks, and encouraged to question, refine, or outright refute, if that is the right thing to do.
6. The objective of the workshop is to make specific recommendations on adoption of a framework for Human Performance and Human Systems Integration for DoD.

Lt Col Valerie Martindale, USAF, presented the challenge, outlining the problems that she has encountered in the fields of Human Performance and Human Systems Integration that can be addressed by a conceptual framework.

1. Organizing the complexity of the human and the task rather than focusing on single or few measures
2. Avoiding dependence on correlation as a method of data analysis
3. Setting meaningful baselines and objectives for performance measures
4. Combining physical, cognitive, and other human resources for performance
5. Determining targets for enhancement such that enhancement produces improved task performance for complex tasks
6. Providing a model which, once populated, can be used to create "performance prescriptions" for a given human in a given situation, for use by a non-expert

She also outlined uses for a framework in determining science and technology investments, crafting policy, selecting, training, promoting, and caring for the operational force, seeking consultation as individuals and small units, speaking to those who write requirements and who conduct human systems integration activities, speaking to engineers about the demands and limitations of acquired systems, and facilitating meaningful communication among all the communities that would benefit from understanding human performance.

She concluded with the objectives of the workshop:

1. Evaluate strategic frameworks for representing human performance.
2. Debate and record pros and cons of frameworks.
3. Derive and record specific recommendations on adoption of a framework to the HSI task force.
4. Make specific recommendations on the way forward to achieve the benefits of a human performance strategic framework for DoD.

Dr George Kondraske, Professor of Electrical Engineering and Bioengineering at the University of Texas, Arlington, and Director of the Human Performance Institute, then presented a brief overview of the reference framework, GSPT and NCRA. He began by pointing out that DoD has a unique ability to look across complex tasks and large populations to synthesize a human performance framework independent of subject matter and agnostic to the origin of the framework.

Dr Kondraske's interest in human performance began with trying to "measure everything" (i.e. thinking about the "total human"), particularly quantifying human neurologic function. This quickly turned to an interest in how measurements are used in prediction of performance on complex tasks, which in turn led to interest in an organizing theoretical framework. He found regression did not work well to derive predictions of performance on complex tasks, such as driving. In addition, performance of complex tasks was itself not well measured or characterized.

He worked at relating a system (which could be a human) to a task, to answer the questions: what can the system do or not do, how well can it do that, and if it can't improve, why not?

Fundamental concepts:

Considering the human as a system which performs a complex task, the human brings resources, called "performance resources," to the task which must all be present in sufficient quantity to accomplish that task, i.e. they exhibit threshold effects. A complex task will draw on multiple resources. Having more of an individual basic performance resource does not improve performance in a complex task when some other resource is rate limiting. A specific amount (threshold) of each resource is necessary but not by itself sufficient.

There are benefits to measuring resources in a manner such that a larger number always reflects "better," i.e. greater resource availability. For example, "reaction speed" is a better term for a performance resource than "reaction time." This has consequences for the predictive ability of NCRA which will be explored later. It also makes communication on the complex task easier by introducing a convention that can be understood by all practitioners.

The Elemental Resource Model (ERM) was then described. ERM is analogous to a "chemistry" for human performance, like baking a cake, where each resource is necessary and none can be related to the final product by an isolated regression. Furthermore, these resources can be specified as a finite set, like the periodic table of the elements. The elements of human performance represent the performance resources that humans combine to produce an infinite number of complex task behaviors, as chemical elements are combined to produce an infinite number of compounds.

Discussion:

Q – Any framework must also take into account the environment and the equipment that the human is working with. An example is Soldier Load, which includes multiple systems, communications, an information load, and protective gear. How are such interactions captured?

A – The system-and-task is a function of conditions. By analogy, in electronics there is a data sheet for transistors. The data sheet includes data at different temperatures. The same concept should apply to humans in different environments. It adds complexity, but does not change the basic model. One problem now is that there is no agreement on core, lower level performance capacity measures, but if a standard existed, people would be willing to invest in obtaining data for more varied conditions (e.g. environmental, etc.). Also, if one understands and models numerically how performance on a higher level task changes when one of these lower level measures changes, sometimes as a function of conditions, one may then predict performance for every task for which a model exists without actually doing a new experiment.

Q – What about higher or lower level systems, human plus machine, or a team of humans? How can such a framework be used to model humans?

A – GSPT/NCRA applies to any hierarchical level. If the F-16 in flight is a system, the pilot is a subsystem. The F-16 squadron could be the system, and each F-16 is a subsystem, and so on. Likewise, within the human, vision could be a subsystem, or vision could be the system and the retina is a subsystem. "Left gripper" and "right gripper" systems could be both hands from one person, or one hand from each of two different people performing a task.

Q –How are subjective vs objective qualities handled? For example, how is motivation handled?

A – Motivation is considered a performance resource of the system. Both subjective and objective measures can be combined in the same framework. When higher resolution measurement is available, it can be used, but binary things can also be incorporated in GSPT as resources. Subjective measurements can be used just as they are now, such as visual analog scales or other subjective measures of a performance resource.

Comment – The Army has an ongoing assessment research project being applied to all recruits that measures motivation. We use it as a predictor of a recruit completing the first term of service. If the recruit scores low, we do not accept them. It is called Tailored Adaptive Personality Assessment System (TAPAS).

Q – Is stress, or ability to cope with stress, a resource?

A –Ability to handle stress is part of the "software" resource of the human. It's built up with training and experience, like adding software to the information handling resources. This is analogous to a laptop cooling fan turning on due to a software algorithm in order to deal with thermal stress. The key here is that a "software" resource still fits within the core resource construct.

David Fitts, Chief of the Habitability and Human Factors Branch in the Habitability and Environmental Factors Division, Johnson Space Center (JSC), NASA, presented the NASA views on human performance and HSI. The Health and Medical Technical Authority is roughly equivalent to HSI for NASA. NASA does not have the mandate that DoD has in the form of DODI 5000.02, to conduct HSI. However, it is developing an HSI culture. The Constellation program included a strong set of contractually binding requirements with respect to HSI. A new Human Health and Performance Center was formed in October that may become a focal point and opportunity to integrate efforts across NASA. NASA is currently willing and able to work on internal change. The cancelation of Constellation and the fast-approaching end of the Space Shuttle era, combined with budget concerns, creates a window of opportunity to review the importance of considering human integration when systems are first envisioned and requirements set. For example, if continued the Orion capsule project provides a platform for reviewing and improving current HSI practices.

NASA standards (notably NASA STD 3001, "Space Flight Human Systems Standards") are well written and lead to contractually binding requirements such as the "Constellation Program Human Systems Integration Requirements." HSI needs to become more active in acquisition of unmanned systems, not just human spaceflight programs, and needs to apply to the entire workforce—i.e. ground support crews, ground flight controllers, maintainers, logistics, and manufacturing. HSI in NASA still faces some resistance from engineers who see it as outside their responsibilities, or as imposing unnecessary constraints on them. The argument for return on investment of Human Systems Integration is not always accepted by decision makers and needs to be reinforced with strong examples and metrics.

Discussion:

Q – How does NASA bridge the engineering/medical culture divide?

A – JSC's Space Life Sciences Directorate is medically (human health) oriented, though discipline strength in and recognition of the importance of human performance is growing. Since program management practices in NASA are built around Systems Engineering principles, to be effective in acquisitions and development, both human health and performance must speak in the language of the engineering community, i.e. solid, objective, measureable SE requirements. It would help bridge any divide for the advertised cost savings from HSI to be backed with objective measurements and examples, but NASA has not yet found the right data and expertise to articulate the cost savings case. The new NASA Human Health and Performance Center could have a role in metrics development and requirements definition in the future. The Center could become a focal point for metrics of human performance.

Q – In DoD there are proven ideas that are not implemented (e.g. Automated Predictive Ground Collision Avoidance System in the Joint Strike Fighter). In the Air Force, traditionally the human is not dealt with directly by the program office; they only consider the interface. We want to deliver to the program manager a tool to deal with human performance and forecasting that supports cost savings. Operational costs are not rolled into development costs, so money is saved in the short run and wasted in the long run. How does NASA get around these hurdles?

A – Like the DoD, NASA has much forward work to perform to facilitate inclusion of operational performance metrics into the development cycle. Once metrics are resolved and agreed upon, NASA has the benefit of being a smaller organization. The HSI office can go to the program office and speak directly to the program manager, can go to their meetings and understand their challenges. Inclusion can still be difficult because of ingrained perceptions among engineers and program managers that HSI can be excluded for cost savings, without regard to resulting increases in the potential for and the cost of mishaps.

Comment – Program managers will take notice if we can give them trusted measures that forecast human performance.

Captain Kris Belland, Pacific Command Surgeon, USN, spoke of his experience in the Navy using the Human Factors Analysis and Classification System (HFACS) to analyze mishaps and derive actionable information on their causes. He analyzed the effects of actions taken as a result of HFACS analysis of operations at Fallon Naval Air Station, Nevada. Results were published (Belland, KM, Olsen, C, and Lawry, R, 2010 "Carrier Air Wing Mishap Reduction Using a Human Factors Classification System and Risk Management" Aviation, Space, and Environmental Medicine Vol 81 issue 11, pp 1028-1032.)

HFACS provides a way to categorize human errors that can provide insight and direct improvements. Actions taken as a result of the analysis focused mainly on supervisory and cultural issues, procedures, and training, not on hardware solutions. Results were demonstrable, showing the cost savings of implementing changes.

At this point HFACS is primarily an analytical retrospective tool; it needs to become proactive. It can provide a data source of performance resources for consideration in the GSPT/NCRA framework.

Discussion:

Q – Can HFACS identify limiting resources in mishap analysis?

A – Any human performance practitioner should be able to do an HFACS analysis and identify what are called contributing factors, which may correspond to limiting resources in GSPT. HFACS looks at four levels: the individual human, management, organization, and culture. Any of those could be the "system" in GSPT. 80% of human factors that cause mishaps are cognitive. This drives home the point that savings are made by incorporating human performance considerations in all we do. HFACS is also being used to investigate medical mishaps and automobile accidents.

Q – HFACS has proven too complicated to get ideas across to decision makers. Instead we find we have to be able to talk about cost every time, to get people to spend the money up front. We need a marketing plan.

Comment – If the science is not clear, the picture to others is not clear. Can you get the operator to take action based on your analysis? They did in the NAS Fallon case, with statistically significant findings. This shows that educating the audience is worth the time.

Comment – We can lose people with complexity if it's not in a trusted framework. We accept complexity in other fields, like medicine, because we have the frameworks of chemistry, anatomy, and physiology. HFACS is successful here (NAS Fallon) because it is systematic, but it still lacks the benefit of a framework for characterizing human systems from a performance perspective, and explaining the nature of the interface of the human system to complex tasks..

Comment – See the Institute of Medicine reports on medical mishaps, To Err is Human (1999) and Nothing has Changed (2010) for the role of human performance in medicine and health care. We need to be able to organize these complex fields of human performance (aviation and medicine) if we are going to make progress in the field, in real world terms, and in speaking to decision makers.

Col Steven Chandler, Chief of the Human Dimension Task Force, Requirements Integration Directorate, USA, presented information on the Army's Human Dimension program. The Army Training and Doctrine Command (TRADOC) commissioned a study resulting in a 250 page report. The conclusion was that the Army needs to learn and organize to handle the "skin in" human as well as it does weapon systems. No weapon system is bought without validating requirements; the same should be done with recruiting the human.

The approach to the total human is characterized with cognitive, physical, and social aspects. Cognitive goals include predicting leadership potential, identifying cognitive styles, and accelerating training. Physical goals include a global assessment tool and a master resiliency program. Social goals cover the character of the soldier, including moral values and professional ethic.

The Human Dimension is applying the Joint Capabilities Integrated Development System (JCIDS) (normally recognized as materiel focused) to validate human capability requirements. The Integrated Capability Development Team (ICDT) is a group of individuals selected to supply critical capabilities to a team, like a JCIDS High Performance Team to write requirements. The JCIDS method was chosen to look at scientifically validated requirements.

One product will be created and validated human capability requirements. Only 27% of applicants are accepted by the Army. 31% of Army recruits are lost. The number one reason for attrition is body breakdown (physical domain). A 5% reduction in attrition would save $108M per year. Reducing attrition and improving performance depend on the same thing: identifying the right combination of individual and task.

One cognitive program is to “train the untrainable,” turning failures into successes by studying the highest performers and teaching their techniques to others. An example is marksmanship. Unacceptable performers can be brought up to an acceptable level, and acceptable performers can be improved beyond their former baseline with this technique.

The number one gap identified by the Human Dimension Task Force at this point is the lack of valid metrics and assessments for performance and for individuals. Many of the assessment tools are copyrighted and require paying a fee for each use. For cost reasons, the Army would like to develop a new set of measurement tools.

Discussion:

Comment – With respect to human capability requirements, there is a mandate in the Navy to create a “sailor system spec” when the Navy buys a new vessel. It’s essential to get the human aspect in early. We need a human readiness level, not just a technical readiness level.

Q – Does the Army bin similar Military Occupation Specialties (MOS)? Are there assessment brackets?

A – Assessments are generic.

Comment – New assessment tools need to be DoD owned, not copyrighted. Assessment information is often proprietary, guarded and hard to get to for a variety of reasons. DoD needs to order sharing of information.

Comment – GSPT uses, and needs, standardized metrics, not more metrics developed in isolation. Open source software should be the analogy. There will be no progress without standardization of measures. Regression is not the appropriate model to use for performance in complex tasks, due to the nonlinear nature of performance resource availability.

Comment – The Marine Immersive Trainer cost $70M, was meant to inoculate against stress, but does not produce stress. In other examples, students have been successfully tested to perform a task in a training situation, but cannot perform in the field.

A – Recently there is not a focus on the soldier himself; rather there is too much focus on process rather than measuring the product of training. We want to “push the benefit curve to the right,” to get more benefit out of people in the organization early, by improving selection, training and assignments (experience).

Moderated Discussion on understanding the many facets of the problem:

A framework needs to provide the context for making trade-offs. For example, if the human cannot adequately supply a resource like night vision, an external system can. An exoskeleton can be considered within the context of the framework. Such enhancements may be appropriate for some tasks but not others. Examples: The human may be expected to integrate hundreds of sensor inputs. An invisible suit does no good if the weapon has a large signature. A framework must provide for these human-materiel and materiel-materiel interactions.

Human performance modeling is dependent on having data at the right level of hierarchy and the right granularity. Currently there is little or no data suitable for modeling human performance. Modeling continues anyway, because we need it, but it is based on assumptions. Data for modeling has to be able to roll up to the task level.

HFACS data has limitations. It is retrospective. It is focused on failure modes, not success modes. It may contain inaccuracy (self-reporting is suspect in investigations). There are difficulties with the coding of errors even by experts that limit its reproducibility.

What about long term consequences, such as the addictive properties of drugs? If we use modafinil and don’t know the long term consequences, how does the framework help? One answer is that we should look higher up in the hierarchy, asking why the task is designed to need the use of drugs in the first place.

Having repeatable, objective, validated information is the way to overcome bias and tradition.

Dr Kondraske presented key concepts in GSPT and NCRA.

The system (human) possesses performance resources, and the task makes demands. Separate the system from the task and employ the resource construct for each and every unique quality that characterizes how well a system executes its function. Resource economic principles explain the system-task interface, such that a threshold level of each resource is necessary to accomplish the task.

A resource is something desirable, and should always be measured in the positive, such that a larger numerical value corresponds to “better.” Measure accuracy, not error rate, for example. A larger number then always represents more resource availability. Resources may not be unrelated, such as flexor strength (torque production capacity) and movement speed, but they are still unique in that they cannot substitute for one another in a certain task, like turning a wrench. To help understand basic concepts, a system with only one dimension of performance can be considered the simplest system. Its resource sufficiency is tested by stressing the system along that dimension. The idea that complex performance is made up of simple elements of performance is called monadology. Each element of performance represents a defined functional unit or system and one of its dimensions of performance that can be measured.

When resources are added, they are combined in the chemical sense, not added in the mathematical sense. They form dimensions along which measures can be multiplied to get an envelope, the Performance Capacity Envelope (PCE). A given task can be either inside or outside the envelope. Successful performance is defined as a task within the envelope. Failure can be outside the envelope along any dimension, or combination of dimensions. Insufficiency in a single resource can cause failure, regardless of the sufficiency of other resources. This is the limiting resource, like the limiting reagent in a chemical reaction. Performance improvement comes from improving the availability of the limiting resource. A nonlinear threshold type of relationship explains the interface between the systems and tasks (performance resource availability must exceed demand imposed by tasks).

Sometimes for a complex task, only expert opinion is available as a measure of performance. GSPT can be applied at any hierarchical level, but the measures used must be appropriate to the level. Systems can be grouped, e.g. all neuromuscular subsystems acting about body joints have the same dimensions of performance, such as torque production capacity, speed, etc. A limiting resource, once identified, can be broken down to the next hierarchical level to identify the contributing weakness.

Any dimension not taken into account is equivalent to saying there is an infinite amount of the resource represented by that dimension. Using regression to relate a basic performance resource to complex task performance does not make sense, because it doesn’t take into account the threshold idea for each performance resource in a combination (vice a mathematical addition) of multiple performance resources.

Discussion:

Q – Can we predict performance from one test score?

A – No, like predicting how much cake you can make from how much flour you have. There will be other resources needed. There are ways to derive composite scores that make sense, however. We are discussing the science of performance theory, but then there is the engineering. We have examples of using NCRA, which we will discuss later, to predict performance well in complex tasks such as driving after measuring 20 to 25 performance resources from a given individual subject over about a 45 minute period. Later we will look at how to form composite scores that are better predictors of performance, but their predictive value depends on using a range of types of performance resource measurements.

Q – Can anatomical limits be considered in this framework?

A – Yes, this gives one tools to establish and account for such limits.

Q – How does motivation come into play?

A – Motivation is a resource. It is multiplied by other resources to give a score – a composite performance capacity. If I have zero motivation, but a lot of strength, I still have zero performance. There are other common examples of multiplying resources, such as for the quantity of work which is force times displacement, and cardiac output which is heart rate or speed times stroke volume. There is also the tradeoff between speed and accuracy (Fitts' Law).

Q – Everyone in an organization has different motives. Do we use a single framework for everyone?

A – If motivation is identified as a limiting resource, the task can be broken down into lower level resources, like spiritual, social, and others, to identify the underlying weakness. If it's not a limiting resource, it can be left at the higher hierarchical level.

Q – How do we deal with resources that are not independent?

A – Can one resource compensate for another? For the space shuttle Challenger, could the O-rings have been compensated for by computers? No. In some cases resources may not be independent, but can still be unique. In some populations there could be a correlation between grip strength and lower limb extensor strength, but they do not accomplish the same thing. They are related but unique resources. In those cases when resources are truly interchangeable, in GSPT using resource economics, they can be described using an "or" function rather than an "and" function.

Comment – The GSPT framework allows comparison among different approaches to performance measurement and prediction.

Comment – GSPT can provide a framework for looking at traumatic brain injury that identifies performance limiting resources.

A – We can look at old studies (datasets) and get new information. Regression never reveals what the limiting resource is in task performance.

Comment – GSPT applies to hybrid, mixed human-artificial systems. The margin of excess resource available is a margin of safety.

A – "Stress" can be defined as the ratio of a resource available to the demand made on that resource. The lower that ratio, the more resource is available beyond demands, the more effective performance can be, and the safer the task or system is.

Comment – The services cannot agree on safety because there is no overarching framework in which to discuss findings and their relevance.

A – We can get agreement on a framework. It is a toolbox from which we can select different tools for different jobs. We can use it at different hierarchical levels, and look at different dimensions without losing the context of the whole. You may not use all the tools for a given purpose, but they're all there in a common toolbox.

Q – How do you see GSPT/NCRA in relationship to the Army's Cognitive Workload Modeling Tool, IMPRINT?

A – They will be complimentary. GSPT provides a framework to support and justify some of the methodology currently incorporated in IMPRINT. It also suggests some new ideas that could easily be incorporated into IMPRINT, which is an evolving tool. The relevance to other aspects of IMPRINT beyond cognitive workload models is similar. I could foresee integration of NCRA approaches as part of this toolbox.

Moderated Discussion on criteria for a framework:

The framework must be objective and repeatable, in the sense that different people using it will come up with same results. It must be usable as a selection tool, for example in medical standards. How should color vision standards be set for performance? What makes a good RPA pilot? How do we avoid another Fort Hood shooting?

The framework has to provide a representation that can be used in modeling. It must have predictive ability. It must be able to be externally validated.

The framework must talk to the acquisition community and to managers. The language and methods must be sufficiently clear and simple and the return on investment sufficiently obvious that acquisition personnel will elect to incorporate the framework.

The framework must be trusted. Standardized language helps. Creation from outside DoD is a plus, i.e. non-political.

The framework must allow intelligent selection of research areas to inform research portfolios and strategies.

The framework needs standardization of language and of measures.

The framework must be operationally appropriate and valid.

The framework must be scalable and hierarchical.

The framework must inform design standards, and must inform capabilities, for example in Development Planning.

The framework must be in simple language, usable by a non-expert. It must accomplish something – solve a problem for the DoD such as attrition, selection, etc.

The framework should be able to accommodate sources of data such as HFACS, ASVAB, etc.

Dr Kondraske presented key concepts in development of composite performance measures using GSPT/NCRA.

To simplify a set of resource measurements into a single number, GSPT indicates that the appropriate operation is to multiply them, not average or add them. To average or add violates the basic principle that “apples and oranges” cannot be added, because they are unique and not interchangeable. Multiplying preserves the units and dimensionality of the resources involved.

Comment – Resilience is hard to measure, may need to be measured in battlefield conditions.

A – Resilience can be a performance resource. Multiplying resilience and physical capacity yields “useful performance capacity.”

The correlation coefficient can be interpreted as indicating how much variability in the data is accounted for by the correlation used. There are some very poor r^2 values (0.41 for some ASVAB-based predictions) being accepted as indicators of human performance. The tendency is to look for better measures instead of a better framework for the measures already identified. Adding or averaging – which is involved in the modeling that has been commonly used to date, -- produces worse predictive results. The concept of multiplication comes from the area or volume of the performance capacity envelope, and improves predictive ability.

Examples from Parkinson’s disease severity – represented by the impact on overall performance -- and upper extremity motion quality show remarkable improvement in the sensitivity and validity of composite scores achieved by multiplication instead of addition or averaging.

Comment – ASVAB is a collection of tests with very poor predictive capability.

A – Maybe part of the problem is the composite score, which is achieved by addition-based combinations of subscores. There will be more information tomorrow when we will go into NCRA and look at how the same data can be analyzed differently.

Dr Bob Ward, Sports Science Network, former Strength and Conditioning Coach, Dallas Cowboys, presented information from his experience in training elite athletes. Using playing professional football as the complex task, he described collecting and analyzing large amounts of performance data. He called his approach, based on GSPT concepts, "the science of enough." A team can be considered a system in GSPT. The team can lose even if it is stronger than the opponent if it is lacking in any one dimension. In studies done by the Cowboys, regression analysis accounted for only 20% to 48% of variability seen when comparing predictive measures to actual game performance.

Factors studied included strength, body fat percentage, speed, agility, power, and performance on simpler tasks such as running, changing direction, and vertical jump. Neurologic function must be included, not just physical function. He had enormous amounts of data from all these measures, and no good predictors using the traditional correlation type of thinking. However, when GSPT was applied to similar data in a quarterback study, prediction grades were increased to within 1 grade point with an accuracy of 81%. In another study that included all 12 football positions, prediction accuracy increased to 84% using GSPT. GSPT changed the analysis to one of resource sufficiency, and made sense of the data.

Q – What about all the intangibles?

A – "Performance" includes all the things you want. There is nothing intangible in performance. Resources may be hard to measure, but they are not intangible.

Comment – No one in pro football is using this because they would be out of a job; it can be used by a non-expert. Experts use personal judgment, not regression models.

Comment – Consider the similarity between performance of a professional athlete in a game and performance of an airplane pilot in a flight. It's difficult to look at all the factors that contribute to performance. There's too much data and no good predictors.

Q – What about overtraining? There's a staleness or plateau. What causes that?

A – There are many types of fatigue; it can be emotional fatigue.

Comment – Coping skills provide resilience. Risk taking can help avoid monotony. What resources could be involved? It could be related to the dopamine reward system.

Comment – It is possible to increase performance from a plateau by getting out of patterns. Varying the approach can increase the envelope.

Comment – Expert task performance is a product of using the individual's resources in the best way, and distributing the task across available resources. The effect of training, especially going from novice to expert, may be thought of as learning how to distribute a task across resources.

Comment – At the team level, you may have to deal with bad leadership. This is relevant in the military context; you could use the concept with military units. Leadership is a resource. Leaders make use of their own resources at the next lower hierarchical level. If leadership is the limiting resource, you can then break that down.

Conclusion: GSPT has provided the best insight out of all statistical models I have used, deriving the most meaningful predictive relationships. You don't need to have the best player to win the game – you need to have the player that is good enough in every performance resource needed to win the game.

---

14 December 2010

Dr Kondraske presented "NCRA and Traditional Alternatives: Developing Performance Models, Predicting Complex Task Performance and Limiting Performance Resources."

When considering a human in a complex task, it is better to measure a broad range of resources demanded by the task than to focus on measuring one or few specific resources with high resolution. It is also necessary to distinguish between measurement, which results in a number (5 ft tall) and an assessment, which results in a judgment (too short to play basketball).

Graph the data with the resource increasing on the y axis and the complex task performance increasing on the x axis (bigger number corresponds to more desirable state). The distribution of scatter plot data when graphed this way is very distinctive, with a distribution that is empty in the lower right corner. Data in the lower left corresponds to low resource availability and low task performance, which makes sense. Data in the upper right corresponds to high resource availability and high task performance, which makes sense. Data in the upper left corresponds to high resource availability but low task performance. This is expected if other factors limit performance besides the resource being measured. For a complex task, data points are expected here. Data in the lower right corresponds to low resource availability but high task performance. The interpretation if data is found here is that the resource being measured is not important to the task.

Using this graphing method, the threshold for acceptable performance can be set on the x axis (based on the data distribution, but how exactly a threshold curve is defined does involve some judgment), and the resource availability needed can be read off the y axis (a data-driven assessment). Measuring a set of resources that are needed for a task and using them with a set of these resource demand functions results in a prediction of task performance capacity.

A single example of a successful human can set an initial standard for a given task across all the resources that can be identified and measured. Measurement of one successful individual's resources can set the upper limit for the thresholds, the least amount needed of each resource to be successful in that task. Each additional successful individual measured cannot raise the thresholds for the set of performance resources considered, but can potentially lower it, by providing an example of someone who could successfully complete the task with less of a given resource.

A task that is inside the performance capacity envelope but near the edge entails risk of failure if the task runs into unexpected circumstances or if the human is exposed to an additional drain on resources (fatigue, stress, unexpected problems).

Measurement can be at different hierarchical levels and granularity depending on the purpose. Basic resources, as in the Elemental Resource Model, should always be in mind. Start with a successful task – the rating indicating success may be a judgment from the commander – and then work down to more basic resources.

Q – The slide shows complex task performance divided into thirds (poor, satisfactory, excellent). Where does that come from?

A – That's just an example. The acceptable level of performance can be determined in any way that makes sense in context. It can be binary. It can be any resolution of breakdown on the scale.

These curves are seen in many different contexts. Examples are given from medical rehabilitation, dementia and driving skills, sports, and neurology. There is a mindset that basic lower level measures do not correlate well with measures at the functional level – and that is a "bad thing." This is actually to be expected, because many resources are in play. When measuring and correlating, one misses the importance of meeting the threshold, and assumes all other resources are available in infinite quantity. Correlation can't be expected with performance data and looking at relationships across hierarchical levels. There is a fundamentally different type of relationship as demonstrated in these scatter plots.

Q – How did someone decide that running quarterbacks through cones was a good test?

A – Tests are based on what you can do repeatably. This was easy to set up and manage. If you can run everyone through similar tests you can get at intrinsic properties. There are tests and datasets already in hand that can be used to make these analyses.

Comment – But we have resources which are functions of other things. Running through cones can be considered a complex task.

A – You can still use the same principles. For assessing quarterbacks you can use that as the basic task, but for research, or for improving a player, maybe you want to break that task down to elemental tasks. Then if you have one quarterback who is really good at everything else, but can't run through cones well, you can break that task down and find out that he's not good at ankle flexion, or some other more basic resource. NCRA is a way to identify the resource that limits performance so it can be targeted for training or other enhancement, or the task can be redesigned to need less of that resource. You start at the higher hierarchical level and work down to find out what you can change, or maybe you can't and that person has to be reassigned. But you know what the limiting factor is.

To use regression it's necessary to have homogeneous populations, with all other variables controlled. To use GSPT, there's no need for population homogeneity. You can work with a realistic and diverse population. You are looking for the threshold amount of a resource that is necessary for success. It may not be sufficient, because there are other resources needed. The performance envelope is set by those who are successful. Those who are not successful can be measured in the same dimensions to find out what resource is limiting to each of them. Which resource predicts performance? The answer is, whichever one is lowest. There is no averaging or other weighting factor.

Q – Isn't this like looking at each resource in a vacuum?

A – That's what is done when looking for correlations between a single basic performance resource and complex task performance. In this case we are measuring multiple dimensions. Only one of them will be performance limiting. If that one resource is bolstered, some other resource becomes performance limiting.

Comment – The way performance predictions are done now, we sometimes don't even recognize or admit there are other dimensions than the one or few being measured. With GSPT, we know there are other dimensions, and we can keep the data on those to use if they become rate limiting under changed circumstances.

A – When using correlation, we are assuming infinite resource availability in all other dimensions. With GSPT, we are saying we know there are other resources needed, we just don't know what the threshold is because they are not performance limiting now.

Q – Can't I just replace one basic resource with another? It seems humans can substitute one resource for another in a complex task, like basic fighter (flying) maneuvers.

Comment – Judgment can compensate for other resources, because judgment is used to distribute the task across resources available. In basic fighter maneuvers, it may be that I know I can't do high G turns, so I conduct the fight in a way that avoids them. That's also why training makes a difference, allowing a complex task to be redistributed across available resources.

Comment – Equipment can compensate for human limitations. You can determine whether the system in your measurement is the human or the human plus the equipment. In the case of terrain avoidance, that task or subtask can be done by the human or by the equipment. If it's done by the human, it becomes a demand made by the equipment.

GSPT/NCRA can be used to identify the current limiting resource for individuals as well as groups, to help them always address that current limiting resource. This avoids wasted training and equipment targeted to something that is not limiting. It also addresses individual differences. Statistical methods only address groups.

Measuring more resources can provide better fidelity of performance predictions. Interaction among resources can be addressed by multiplying them and seeing if that provides a better prediction.

Moderated Discussion:

There is a lot of data that can be reanalyzed to get new insights. Tests used now that seem poor predictors may still be valuable, but we don't know if we don't correctly interpret the findings as resources for complex higher level tasks and take into account the fundamental principles involved.

What DoD datasets exist that could be reanalyzed using the GSPT/NCRA framework?

- Remotely Piloted Aircraft pilot and sensor operator selection
- Navy Fighter Selection
- Undergraduate Flying Training selection and training
- Operational Based Visual Assessment
- Armed Services Vocational Aptitude Battery (ASVAB) and its subtests
- Defense Health Board data on suicides, perhaps POW database for resilience
- Battlefield Airman selection, training, equipment

DoD has huge amounts of data, much of which goes unanalyzed or underanalyzed. In every case, getting access can be very difficult. Reanalyzing this existing data could be advantageous in several ways, particularly in providing new insights by augmenting analyses done by other methods. The analysis is much less expensive than the initial data collection. It's difficult to write grants to reanalyze existing data, however; granting agencies want to see new data even when it's not needed. Also GSPT as a type of analysis is not yet as well-known as, for example, analysis of variance (ANOVA), and therefore has to be explained in detail each time a proposal is written.

Q – How do you decompose a complex task in GSPT?

A – You can use the traditional human factors or HSI approach or any other traditional approach, measuring the same types of things you would have measured for a correlational study, but use GSPT/NCRA analysis. To use GSPT, you always have to think in terms of performance resources and may need to transform some traditional things like reaction time to reaction speed as discussed previously. One advantage of GSPT is that if you don't have all the resources accounted for, the model overpredicts performance. That is, you will find as part of performance model evaluation that differences between model prediction and your gold standard reference for complex task performance are such that the model predicts performance better than the actual performance. This is the clue that there are important resources not yet accounted for.

Comment – It would be good to see the validation of testing and training systems with this approach. For example, is it worth the decision to invest in a color vs a black and white monitor for a training system? It would also be useful for selection criteria validation.

Comment – Training takes up more and more time. It would be nice not to waste time on training that does not help.

Comment – Training can be altered to target different learners. One way is by using learning style. Another is by training targeted to each person's limiting resource.

Q – Can GSPT/NCRA be used on a weapon system?

Comment – It can be used on equipment, training, groups and teams – any system. It can be used to provide data-driven acquisitions with respect to HSI. It can also be used for processes, which are also systems.

A – We already use a performance envelope to describe the capabilities (resources) of the plane; we can just extend that to the human plus the plane.

Comment – We can use this to look at visual acuity and get to minimum requirements that are evidence-based.

Comment – Standards for systems can be set to accommodate available human performance. If standards have to be lowered, for example because of the available recruiting pool, we can predict the impact to total human-system performance.

Dr. Malcolm Stewart, Neurologist, presented, "GSPT, NCRA and Perspectives on Medical Contexts." He began by comparing the medical paradigm with GSPT. The medical paradigm deals with disability, symptoms, and subjective measurements. The GSPT paradigm deals with performance, limiting resources, and objective measurements. The two paradigms are complementary.

He presented data from neurologic examples. Clock drawing is a test frequently used to assess higher level neurologic function, because it draws on many resources. Using this test and NCRA, he analyzed Parkinsonian patients, including one who was revealed by the test to be misdiagnosed. GSPT/NCRA can be combined with path analysis to provide information on which resources are most critical/predictive because of the central or multifunctional role that some resources play in a given task. There is also information in the shape of the resource demand curve, such that the degree of performance improvement elicited by an increase in resource availability can be predicted ("bang for the buck").

It is possible to use subjective measures of resources effectively. Multiplying subscore tests rather than adding or averaging provides a much better correlate to subjective changes in health and ability, and a more valid measure of status changes.

Using GSPT to do evaluations of personnel throughout a career supports maintaining their resources and therefore their performance, predicting changes that affect performance, and targeting interventions. This use complements HSI in safety and acquisition, where demands made by systems on the resources of the human are determined, and the resources for task performance supplied by systems are determined.

Bottom line: Focus on a performance paradigm rather than a disease paradigm. Look at the full spectrum of resources to include physical, social, and cognitive. Select, evaluate, train, monitor, and intervene as needed, to support, enhance, and maintain performance. Design tasks and equipment around the resources of the human.

Discussion:

Comment – In the military we need to establish a performance paradigm for medicine, with capability documents for the human and a Human Test Evaluation Management Plan followed throughout a career. This will decrease lifecycle cost, but the approach is considered somewhat revolutionary.

Comment – We need to be able to articulate to the decision makers, which means general officers, what problem we are solving. Is it attrition, safety, cost, selection, training? What are we trying to solve? What will the tool do for him?

Wayne Ensign, PhD, SPAWAR, presented "Scalability Issues and Broad Perspectives." He addressed fatigue and stress. Stress is a side effect of sleep deprivation. (Sleep deprivation is not the only cause.) The response to stress is more important than the stress itself in determining performance.

Distributed operations in the military results in young individuals with enormous firepower at their command. He observed dismounted Marines at Combined Arms Training and Urban Warfare Training and used measures of life demands and of athletic overtraining. He found there is a point at which no more adaptation takes place, calling into question the value of extended, immersive training.

He also used critical incident reports from combat units to deduce the effects of stress and fatigue on cognition. Situational awareness is often cited as lost, but is difficult to define. Mica Endsley has the most accepted definition, but the three levels may not be independent. Complacency is often cited as a cause of critical incident failure. Does it correspond to loss of the first level of situational awareness, perception, or to a higher level? How does training relate to situational awareness? Lapses in judgment are cited four times more often in operations than in training situations.

Using Mojave Viper training for Marines immediately pre-deployment, his team measured fatigue in 12 units and 9 small unit leaders. Again the effectiveness of the training is questioned because its demands put trainees in a non-adaptive stress zone.

Q – What about the Yerkes-Dodson law that relates stress and performance?

A – It depends on the task and the individual. The emotional activation is key to the response. Without emotional activation, there may be no stress.

The reciprocal of fatigue can be considered an elemental resource (measured so that a larger number is better, representing greater resource availability for the task). Some people are fatigue resistant. This can be measured and can also be considered a resource.

Evidence from f-MRI, cell biology, and biochemical studies supports a mechanism by which sleep deprivation leads to cell death in the CA-1 region of the hippocampus, the area associated with memory. The damage appears reversible. This illustrates the scalability issue that a framework for human performance must be able to accommodate – a basic resource at the level of cell biology has measurable effects on team performance, many hierarchical levels higher. Looking at it in reverse, a deficit in team performance can be traced to more basic, elemental resources.

Sleep can also be a resource. Can we learn to sleep more effectively? We can apply what we know to improve duty cycles and sleep conditions. After days of sleep restriction, Marines were not back to baseline even after three days of unlimited sleep.

Dr Kondraske presented information on predicting complex task performance with NCRA. He presented examples from physical rehabilitation patients (a diverse population) using an appropriate obstacle course and on prediction of laparoscopic surgical ability in medical students. Performance will be over-predicted if an unidentified resource is limiting. This emphasizes the need to measure a range of different types of resources. The resolution of those measurements is not as important as representing different dimensions of performance.

There is also a dimension of time. Demand fluctuates over time, and resource availability fluctuates over time. Some resources, such as endurance or vigilance, include a time dependence concept.

LCDR Joseph Cohn, PhD, Military Deputy, Human and Bioengineered Systems Division, ONR, presented, “Sharpening the Edge: conceiving the challenge and solution space for improving human performance through human systems integration.” He presented a brief history of thought on the human system and human performance to give perspective in understanding where we are today. The current challenge is "to merge the human operator with the technology to achieve overall performance improvement unattainable using either element alone."

We need to describe, characterize, and model the human system (the human mind). Stop treating human performance as "someone else's problem," and allowing those responsible at each step in the process to assume someone else will address it.

Accepted models of human performance have been based on observable behavior, but this only infers cognition. Can we model cognition itself? Can we integrate for example the human and the aircraft to work as one system? Improvements in understanding the human mind will be needed. Some examples were given from current cognitive neuroscience. Measurement tools allow high resolution of data, but solutions are not blended. Most importantly, we want to forecast.

We need to be able to combine different types of data meaningfully. We need to have many tools in the toolset. We need to have a framework on which to hang what we know to see how it fits together.

Comment: This is a good summary restatement of the problem we started with yesterday.

Moderated discussion: evaluate GSPT

Discussion covered the criteria identified for a framework. Findings are summarized here:

1. As much complexity as you can find, it can handle.
2. It can take low or high resolution data.
3. It avoids dependence on correlation for data analysis.
4. It can be used to set meaningful baselines and objectives for human performance.
5. It can be used to set minimum standards that are based on data to achieve a measured level of performance. To get good results, we must incorporate measures that are reasonable.
6. It can be used to write better requirements.
7. It can meaningfully combine different resources, such as cognitive, physical, social, into one model. It can provide meaningful composite scores with better predictive ability than current methods.
8. It can be used to identify and characterize limiting resources.
9. It can handle measurements that are not truly independent, and still give good usable results. This makes the determination of meaningful dimensions of performance less onerous. Example: there is a known tradeoff between speed and accuracy (Fitts' Law), but speed and accuracy can still be measured independently and used as two dimensions of performance, without affecting the validity of the results.

10. It can be used to eliminate assessments that we don't need, that don't measure resources of importance.

Discussion revealed some questions for further work:

1. Can it be modified for use by a non-scientist? Can it be used in the field with an indicator as simple as yes/no, or red-yellow-green? It needs to be a tool that I can trust that is as simple as the dashboard in the car.
2. Who is the consumer? Researchers, modelers, policy makers, requirements writers, engineers, operators? Are we asking for a transformation across communities? Do we need one? We need an educational approach for practitioners.
3. We need to explain to high level decision makers how performance saves them money. We need to be able to present a problem and the solution in terms that decision makers care about and understand immediately. We need the "elevator speech." We need a marketing strategy.
4. We need to be aware of Admiral Mullen's Total Force Fitness. We also need to have taxonomy in line or at least translatable among various human performance programs.
5. Can we ask that new research on performance be analyzed this way?

CONCLUSIONS:

At this point, the group rapidly converged on consensus statements, included in the next section. Discussion centered on obtaining datasets to reanalyze and targeting problems of interest to DoD. Attrition was identified as the number one problem, with selection, that is matching the individual to the task in terms of military specialty, playing the largest role. Additional targets include identifying and making the case for safety improvements that improve performance, improving training (targeting, validating, and setting recurrence rates). Several research uses emerged, to include identifying resilience predictors/resources, and investigating suicide and private motor vehicle safety. Immediate use could be made in setting medical standards in vision. Work on articulating human performance capabilities and limitations to operators for inclusion in requirements, and to engineers for inclusion in design and development, will come later. There is definite interest in this from the services' Human Systems Integration offices. NASA expressed a continued interest in developing the framework and its applications.

An additional opportunity surfaced in that a software prototype for GSPT/NCRA analysis exists. It is not yet ready for beta test. Development of such software for analyzing existing datasets would provide a significant savings in time and money and provide immediate results on the applicability of the theory to a given dataset.

The software prototype is a very general purpose tool for use to do quantitative, performance-based task analysis i.e., to develop performance models for any complex task selected by the user and for which the user has collected suitable data. It can then use a performance model created to predict complex task performance for one or more individuals given only measures of "more basic" or "lower level" subsystem performance capacities. It will also identify which of the lower level capacities is currently the limiting performance resource for a given individual.

As such, it is somewhat analogous, from a user's perspective, to regression analysis software. The data fed in will be a) complex task performance level - like the dependent variable in a regression analysis; and b) "more basic" performance capacities - like the independent variables in a regression analysis - and the result is a model. However the "form" of the model is substantially different; it consists of a set of "resource demand functions" that indicate how much of each basic performance resource is required to achieve a given level of performance in the complex task.

Once the model is built new data can be run through the model to obtain a prediction of complex task performance for each individual one chooses to evaluate. And, again, there is an added benefit (compared to regression analysis) in that the NCRA software will identify which basic performance resource is "the limiting performance resource."

CONSENSUS STATEMENTS

- There is strong consensus that GSPT and NCRA show great promise as a very capable, evidence-based framework with multiple military applications.
- Correct application has the potential to: improve personnel selection; reduce attrition; optimize training; inform research in science and technology; improve / validate standards for acquisition requirements, development, and engineering (HSI); improve performance of individuals, teams, and systems; and thereby reduce costs and mishaps.
- Recommendations:

1.1. Demonstrate the concept by analyzing existing datasets that will rapidly result in actionable information: Remotely Piloted Aircraft operator selection criteria; Armed Services Vocational Aptitude Battery; Battlefield Airman; and Lackland Behavioral Questionnaire.

1.2. Fund NCRA analysis of performance data in selected new studies for direct comparison with traditional data analysis.

1.3. Fund maturation of NCRA software (now in development) for beta test.

1.4. Broaden the understanding of GSPT and NCRA by creating collaborations with:

- 1.4.1. Acquisition Technology Program Task Force
- 1.4.2. Uniformed Services University Total Fitness Program

1.5. Follow up with future meetings:

- 1.5.1. Maintain the current core group
- 1.5.2. Include more operator representation
- 1.5.3. Include acquisition representatives (HSI, systems engineering)
- 1.5.4. Include Health Affairs representation for medical accession standards
- 1.5.5. Hold a telecon in three months
- 1.5.6. Hold another workshop in six months to evaluate progress, evaluate emerging results, and address next steps

Acronyms Used

AFB Air Force Base

ANOVA Analysis of Variance (a statistical method)

ASVAB Armed Services Vocational Aptitude Battery

CA1 This is not an acronym. It refers to a histologically distinct region of the hipplocampal structure in the brain.

DoD Department of Defense

DODI 5000.02 Department of Defense Instruction 5000.02, "Operation of the Defense Acquisition System"

DSOC Defense Safety Oversight Council

ERM Elemental Resource Model

GSPT General Systems Performance Theory

HFACS Human Factors Analysis and Classification System

HPWG Human Performance Working Group (of the Human Systems Integration Task Force)

HSI Human Systems Integration

HSI TF Human Systems Integration Task Force

ICDT Integrated Capability Development Team

IMPRINT Improved Performance Research Integration Tool, a tool for modeling workload in task accomplishment

JCIDS Joint Capabilities Integrated Development System

JSC Johnson Space Center

NAS Fallon Naval Air Station, Fallon, Nevada

NASA National Aeronautics and Space Administration

NASA STD 3001 NASA Standard 3001, "NASA Spaceflight Human System Standard"

NCRA Nonlinear Causal Resource Analysis, a method of analyzing the relationship between task performance and performance resource availability

PCE Performance Capacity Envelope

POW Prisoner of War

RPA Remotely Piloted Aircraft

SPAWAR Space and Naval Warfare Systems Command

TAPAS Tailored Adaptive Personality Assessment System

TF Task Force

TRADOC US Army Training and Doctrine Command

USAF US Air Force

Organizing Committee

Lex Brown, Col, USAF, MD
Director, Human Performance Integration Directorate
711 Human Performance Wing
Air Force Materiel Command
Brooks City-Base, Texas
Lex.Brown@brooks.af.mil

Valerie E. Martindale, Lt Col, USAF, BSC, PhD
Plans and Programs Office
711 Human Performance Wing
Wright-Patterson AFB
Valerie.Martindale@wpafb.af.mil

Anthony P. Tvaryanas, Lt Col, USAF, MD, PhD
Human Performance Integration Directorate
711 Human Performance Wing
Wright-Patterson AFB
Anthony.Tvaryanas@wpafb.af.mil

George V. Kondraske, PhD
Director, Human Performance Institute
Professor, Electrical Engineering and Bioengineering
University of Texas at Arlington
Kondraske@uta.edu